PROBE DATA FROM CONSUMER GPS NAVIGATION DEVICES FOR THE ANALYSIS OF CONTROLLED INTERSECTIONS

Arnold Meijer (corresponding author)
Business Development Specialist, TomTom International
P.O Box 16597, 1001 RB Amsterdam, the Netherlands
Phone: +31 20 7578107, Email: arnold.meijer@tomtom.com

Bart van Arem
Full Professor, Delft University of Technology
Department of Transport and Planning, Faculty of Civil Engineering and Geosciences
P.O. Box 5048, 2628CN Delft, the Netherlands
Phone: +31 15 27 86342, Email: b.vanarem@tudelft.nl

Maria Salomons
Post Doc. Researcher, Delft University of Technology
Department of Transport and Planning, Faculty of Civil Engineering and Geosciences
P.O. Box 5048, 2628CN Delft, the Netherlands
Phone: +31 15 27 88556, Email: a.m.salomons@tudelft.nl

Peter Krootjes
Traffic Engineer, TomTom International
P.O Box 16597, 1001 RB Amsterdam, the Netherlands
Phone: +31 20 7575317, Email: peter.krootjes@tomtom.com

Nick Cohn
Sr. Business Developer, TomTom International
P.O Box 16597, 1001 RB Amsterdam, the Netherlands
Phone: +31 20 7575302, Email: nick.cohn@tomtom.com

Final paper submitted for publication to ITS-America 2012
04-27-2012
ABSTRACT

Probe data from consumer GPS navigation devices provides a network-wide and cost-efficient data source for measuring vehicle movements, whereas experimental studies have been confined to small datasets. Data collection from road-side sensors can provide similar information but is expensive and requires high maintenance expenditures. This paper presents the results of intersection performance measurement using a large probe dataset from consumer GPS navigation devices. The measurement of delay and turning movements is compared with loop detection traffic counts and delay estimation from stationary sensors in the Dutch city of Delft. Turning movements are measured with an average error of 1.35% to 3.76%. The results show that the data is beneficial for a quick assessment of the delay and turning movements at intersections, especially intersections where data from stationary detectors is not easily accessible.

Keywords: Traffic monitoring, probe vehicle data, GPS navigation, intersection control
INTRODUCTION

The measurement of traffic volume, turning movements and delay at intersections is needed for efficient intersection control. Optimization of road geometry, intersection control and dynamic traffic management requires up to date and accurate traffic information. The greater part of intersection performance studies are comprised of evaluation of measurements from road-side sensors and loop detection. Data collection from road-side sensors can directly be used to assess the performance of an intersection but it is expensive and requires high maintenance expenditures (1). Loop detection delivers accurate traffic intensities but can suffer from detection errors and missing data (2). Continuous use of road-side sensors requires constant availability of detection equipment at the intersection, which is often not the case at uncontrolled intersections.

Recent studies using a confined dataset have shown that probe data offers great opportunities for the determination of performance at intersections (3), (4), (5), (6) but needs a certain level of penetration of probe vehicles in the network (7), (8), (9). Probe vehicle data using a large dataset from consumer GPS navigation devices provides an interesting alternative. The use of GPS navigation devices is increasing rapidly (10), providing the momentum needed for a network wide application of probe data in traffic studies.

TomTom, one of the largest manufacturers of consumer GPS navigation devices in the world, has been collecting probe data from GPS navigation equipment since 2007. The size of the dataset is dependent on the anonymous and voluntary contribution of GPS navigation users and comprises location measurements of navigational equipment on a global scale (11). Privacy filtering is used to remove all links to a user. The device identification and trip ends are removed before the data is stored. Map-matching algorithms link the GPS location of vehicles to the road network, producing a network-wide probe dataset.

This paper describes how delay and turning movements are derived from probe data collected with consumer GPS navigation devices at two intersections in the Dutch city of Delft. The results are compared to turning movements from loop detection collected at Regiolab Delft (12) and delay estimation models. We explain our measurement of turning movements from loop detectors and a selection of a time-dependent stochastic delay model. Hereafter we describe the configuration of the test case area and the methodology of probe data collection with consumer GPS navigation devices. Finally, the results of the study are discussed and the conclusions and remarks for further research are presented.
process of arrivals and departures at an intersection, measured manually or automatically. The green-request detection loops at vehicle-actuated intersections are an input source for the traffic control device but also provide the opportunity to conduct traffic counts at the intersection. For this reason loop detectors are often used as input for performance studies at intersections and serve as a reference in this report.

Turning movements at intersections are described by an Origin-Destination (OD) distribution or OD matrix which is derived from traffic counts for every OD combination. The OD distribution is presented as the absolute (traffic volume) or relative (percentage) distribution of turning movements at the intersection.

The delay is defined as the difference between uninterrupted and interrupted travel times through the intersection. Measurement of travel time occurs by measuring the time difference between the arrival and the departure at the intersection. To ensure minimal influence of external factors and driving behavior, the arrival and departure locations should be far enough from the intersection to include the braking and acceleration behavior in the measurement of the travel time. Delay is calculated by comparison of the measured travel time and the free-flow travel time (the travel time if there was no interference on the intersection):

\[ T_{\text{Delay}} = (T_{\text{Departure}} - T_{\text{Arrival}}) - T_{\text{Free-flow}} \]  

(1)

The delay of vehicles on intersections is not directly assessed by traffic counts from loop detectors which, apart from a few experimental cases, only measure traffic volume. Placement of additional measurement equipment such as camera detection and remote sensing can directly measure vehicle delay, but induce much higher investment costs. Traffic counts from loop detectors combined with traffic control device monitoring provide an alternative for expensive measurement systems. In this research study a steady-state approach by Webster (14) and Akçelik (15) is selected for the estimation of delay based on traffic counts from loop detectors and signal monitoring. The average delay at the intersections is calculated on a macroscopic level, to serve as input for LOS assessments. The approximate value of the total delay (delay rate) for a movement at isolated fixed time signals (D in vehicles) is expressed as follows:

\[ D = \frac{q_c(1-u)^2}{2(1-y)} + N_0 y \]  

(2)

Here \( q_c \) is the average number of arrivals per cycle in vehicles, \( u \) is the green time ratio, \( y \) is the flow ratio and \( N_0 \) describes the overflow queue in vehicles. The flow ratio \( y \) is the ratio of the arrival flow and the saturation flow of the movement. Furthermore the average delay per vehicle (\( d \)) is derived (with \( q \) the flow of vehicles per second):

\[ d = \frac{D}{q} \]  

(3)
The described method is developed for isolated fixed-time signals but also provides delay calculation in case of vehicle actuated signals. Measurement of the green activation times, the length of the green phases, the signal sequence and the intensities for each separate cycle provides delay estimation for vehicle actuated signals. An exception for this model is the use of multiple green phases per signal sequence. This case can be included in the first term at Equation (2) and is explained in further detail in (15).

MEASUREMENT SETUP

A field location was selected to evaluate the accuracy of probe data collected with consumer GPS navigation devices for the measurement of delay and turning movements at intersections. The field location comprises two intersections in the Dutch city of Delft. As a reference the traffic volume is measured with loop detection equipment at traffic signals and the green phases are monitored at the traffic control device:

FIGURE 1 Digital map and location of detectors at the test case intersections
Meijer, A., van Arem, B., Salomons, M., Krootjes, P. and Cohn N.

The speed limit at the first intersection is 70km/h on the main stream (East link) and 50km/h on the secondary streams. At the second intersection the speed limit is 70km/h on all links. On average, intersection 1 handles a total traffic volume of 27,000 vehicles per day and intersection 2 handles a total traffic volume of 17,000 vehicles per day. At intersection 1, loop detection is not placed at directions 7, 8 and 12. At intersection 2, loop detection is not available at directions 2 and 3.

**PROBE DATA FROM CONSUMER GPS NAVIGATION DEVICES**

Currently 40% of all vehicles in the United States and Europe are equipped with some kind of GPS navigation (16) but only part of these devices is suitable for probe data collection. Application of analytical sample studies and data enhancement enables the calculation of traffic flow characteristics without equipping all vehicles with probe data collection equipment.

The probe data selected for the research study in this report is part of the Floating Car Database (FCD) collected by TomTom. The probe data is collected from Personal Navigation Devices (PND) under consent of users. The probe data comprises GPS location measurements which are stored on the PND together with the time of the measurements. The GPS receiver stores the location of the device during every second of a trip. For privacy reasons TomTom assigns an anonymous identification code to a device daily to track the vehicle movements and a single probe can only be assessed for a one day period. The collected data comprises on average 300,000 location measurements per day in the city of Delft during the test period of April 2009 until June 2009. The number of measurements, and with it the resulting sample size, is defined by the amount of users of GPS navigation that voluntarily share their data.

The individual measurements are combined and stored as turning movements at the test case intersections based on the time of crossing, the covered distance and GPS accuracy at the time of measurement. Using a multi-source multi-destination Dijkstra algorithm each individual measurement is analyzed and is linked to the most probable location on the road based on the chosen route of the vehicle. The resulting dataset comprises for each turning movement the moment of arrival, the moment of departure and the OD of the movement.

**ANALYSIS OF TURNING MOVEMENTS**

The OD distribution defines the ratio of the total traffic volume at the intersection per turning movement within a specified time period. To evaluate the accuracy of measurement of turning movements derived from consumer GPS navigation devices, the OD distribution is calculated for both intersections. Probe counts are assessed per turning movement and the distribution is calculated for the complete intersection. The results of the probe data distribution are compared to turning movements which are derived from traffic counts conducted with loop detectors. Individual traffic counts from loop detectors combined with
the conservation of vehicles result in the OD distribution of the intersection. This OD distribution is set as “ground truth” and serves as a reference for the measurement of turning movements with probe data from consumer GPS navigation devices.

**Sample size measurement**

The measurement of turning movements starts by defining an optimal sample size i.e., the number of measurements. The sample size is determined by the number of probe measurements that are shared by the user of GPS navigation. The selection size increases with the length of the collection period. Increasing the collection period to an indefinite timeframe is not recommended; longer collection periods lower the level of detail and enable seasonal changes to affect the distribution of turning movements. To determine an optimal collection period for the case study, the error of the turning movements is used. The error is defined as the average absolute error of all individual movements:

\[
Error = \frac{1}{n} \sum_{i=1}^{n} \text{abs} \left( \text{reference}_i - \text{probedata}_i \right)
\]

with \( n \) equaling the number of turning movements.

An optimum is derived by iteratively measuring the total absolute error of the OD distribution measured from probe data compared to the reference situation. After each iteration, the length of the collection period is increased by one day directly increasing the sample size. The resulting absolute error related to the sample size is shown in Figure 2.

![Figure 2](image)

**FIGURE 2** The average error of the distribution of turning movements compared to the sample size of the research study

The total error at both intersections approaches a constant level at approximately 950 measurements. At intersection 1 a stable result is visible after 12 days with a sample size of approximately 900 measurements and for intersection 2 this comprises 8 days with a sample
size of 950 measurements. It appears that the minimal error is obtained if the length of the collection period is fit to the optimal sample size. For smaller time frames (for example morning rush hour) the calculation of the turning movement distribution requires a longer collection period to reach optimal sample size.

**Distribution of error**

To investigate the nature of the measured error, the distribution of turning movements is calculated with a sample size of 950 measurements. Note that turning movements are used only if loop detection information is available, since this data is used as reference.

The results show that a great part of the total error is located at two or three of the turning movements. At the first intersection turning movements 1 and 2 account for 48% of the total error and on the second intersection turning movements 1 and 9 account for 56% of the total error. The results indicate that the error is caused by over- or under-representation of probe vehicles in these movements. We assume that this is caused by possible different travel motives for people who use a GPS navigation device compared to other road users. At the first intersection traffic counts on the deviating movements are conducted with one single detector which may cause an error in the measured traffic counts and result in a shifted distribution. The distribution of traffic and the error are displayed in Figures 4 and 5.

![FIGURE 4 The turning movement distribution for intersection 1 derived from loop detection, probe data and the absolute error per OD movement](image-url)
FIGURE 5 The turning movement distribution for intersection 2 derived from loop detection, probe data and the absolute error per OD movement

ANALYSIS OF DELAY

The measurement of delay from consumer GPS navigation devices is evaluated by using a comparison with results from a time-dependent stochastic delay model, traffic counts and signal monitoring. In the case study no ground truth reference is available for comparison. The time-dependent stochastic delay model provides an estimation of the average delay and acts as a guideline to analyze the credibility of the results. The delay distribution is calculated for separate movements and the results are explained with examples.

Measurement approach

With probe data the travel time of vehicles is directly measured. Travel times of probe vehicles are measured by comparing the difference between the passage times upstream and downstream of the intersection. For all probe vehicles that enter the study area the turning movement is identified and the related travel time is stored. The delay is defined as the difference between uninterrupted and interrupted travel times through the intersection. For this knowledge about the free flow travel time of a movement is required. A method is proposed for the measurement of the free flow travel times at the intersection. Based on the local speed limits a reference free flow travel time is calculated. In an iterative process the lowest measured travel times are identified as free flow movements within a 95% probability interval of the reference free flow travel time. This comes down to an average of 2% of all intersection crossings. The average travel time of the free flow turning movements is selected as the free flow travel time.
The proposed method ensures that a maximum sample size is used for the calculation of the free flow travel time increasing the reliability of the measurement. An advantage of this method is that the calculated free flow travel time is actually measured and is fit to a specific intersection and movement. Factors which influence the free flow travel time such as the layout and surroundings of the intersection are automatically taken into account. A downside of the proposed method is that the calculated free flow travel times already take into account the delay which is caused by the signal control device.

Implementation of the free flow travel time in Equation (1) provides the delay per probe vehicle turning movement. The individual probe vehicle delays from the complete three month period are combined to create the average delay distribution per day. The average delay is calculated for 15 minute aggregation intervals. A short analysis shows that 15-minute intervals provide more granularity than one-hour intervals and for smaller intervals do not provide complete data coverage. The average delay of the probe vehicle data is compared to the delay distribution derived from the time-dependent stochastic delay model. This is done between 07:00 and 21:00. For the analysis, only working days are taken into account to create a clear picture of the differences between rush hour and off peak conditions.

**Comparison of results**

The delay distribution shows a similar trend in the results of the time-dependent stochastic delay model, but with more noise in the distribution. A logical explanation for the greater variance compared to the time-dependent stochastic delay model is that the sample size in the three month period, 7,262 observations at intersection 1 and 12,980 observations at intersection 2, is too low to create a smooth distribution. Testing with smaller sample sizes confirms this hypothesis. The analysis is confined by the length of the collection period and cannot be increased to increase the sample size. To demonstrate this result a representative movement is selected and the average delay distribution is calculated. Figure 6 shows the results for left turning movement 4 at intersection 1.

![Comparison of the average delay per 15-minute time period for left turning OD movement 3 at intersection 1 between 07:00 and 21:00](image)
Although the results in the example show a great variance, the overall average delay shows a similar trend to the time-dependent stochastic delay model. The model measures an overall average delay of 37.6 seconds compared to the delay of 38.1 seconds measured with the probe data, a difference of 1.3%. In order to clarify the comparison, we propose to apply a method to reduce the amount of noise and locate trends in the distribution. A moving average filter and a Savitzky-Golay filter are selected for the smoothing of the probe delay distribution. Clearly the smoothing process results in a reduction of noise and the probe distribution shows more resemblance to the results of the time-dependent stochastic delay model. The trends in the distribution are clearly visible and the variance is lowered. However the smoothing process also reduces peaks in the distribution resulting in data loss. In the example this results in a maximum peak reduction of 35% for the Savitzky-Golay filter and 60% for the moving average filter (Figure 7).

The greatest deviation between the probe delay distribution and the reference data is measured during rush hour periods. During these periods probe data shows substantially higher or lower delays than the time-dependent stochastic delay model, which can lead to differences of up to 20 seconds. It is not possible to locate the source of this deviation, the measured deviation is different for each movement and both positive and negative deviations are measured.

The calculated free flow travel time takes into account the delay caused by the signal control device. The results show the effect of this assumption. At periods with a very low traffic demand (night time) the delay distribution approaches zero seconds on some of the movements. The time-dependent stochastic delay model however indicates a delay of 10 to 15 seconds during these periods. Figure 8 shows an example of this trend which starts at approximately 19:00 for movement 11 at intersection 2. It is debatable how this effect influences the distribution before 19:00.
To investigate the variance of the average delay, the distribution of the delay measurements from the individual probe measurements is calculated. We compare the sample of probe vehicles with the accounted delays at a single movement. The comparison shows a random arrival pattern for the collected probe data. The accounted delay of the probe vehicles appears to be normally distributed, most likely caused by a random distribution of probes amongst traffic. We derive from the results that the error of fit decreases if the number of probe observations is increased. Figure 9 shows an example of this distribution for left turning movement 9 at intersection 2 with $\mu=43.7$ and $\sigma=23.3$. 

FIGURE 8 Application of a moving average filter to the average delay per 15-minute time period for through OD movement 11 at intersection 2 between 07:00 and 21:00

FIGURE 9 Distribution of delay amongst observations for left turning OD movement 9 at intersection 2
DISCUSSION

The study shows that a minimal number of probe measurements are required for the measurement of delay and turning movements at intersections. For practical use in LOS assessments the number of probe measurements defines the length of the analysis period, resulting in intersection performance on a per-minute level. For traffic studies on a per-second level an increased amount of probe data from consumer GPS navigation devices would be required.

The selected time-dependent stochastic delay model provides an estimation of delay but actual measurement is required to evaluate the real strength of the delay measurements. The question arises if the model by Akçelik (15) is the correct model to serve as a reference for the evaluation of probe data which measures travel times compared to the estimation of delay. Because of the nature and quality of the probe data it is questionable which dataset should serve as the reference for delay measurement. It is advised to investigate the measurement of delay with travel time measurements from accurate reference systems such as camera recognition to clearly define the error of measurement. Furthermore it should be evaluated how the model by Akçelik (15) should be adjusted when applying it for probe data analysis.

In the study two intersections are examined. Additional effort is advised to investigate the accuracy of the probe data at intersections with different layouts and under deviating traffic conditions. The effect of the number of probe measurements and the penetration of probe vehicles should be analyzed at multiple intersections to check the network-wide representation of the data.

CONCLUSIONS

This paper analyzed probe data from consumer GPS navigation devices for performance measurement at intersections. The paper compared the delay and turning movements collected with consumer GPS navigation devices at two intersections in the Dutch city of Delft with measurements from loop detectors and traffic signals. The measurement of turning movements was analyzed with a ground truth reference from loop detection traffic counts providing a clear definition of the measured error at the intersections. The delay measurement was compared to a time-dependent stochastic delay model using input from loop detection traffic counts and signal monitoring.

The study indicates that probe data from consumer GPS navigation devices provides a data source for measurement of delay and turning movements at intersections. The system satisfies an average penetration rate of 0.5% and applies an update frequency of 1 measurement per second. Under these conditions turning movements are determined with an average error per movement of 1.3 to 3.8%, mainly caused by the error at two or three movements. At the test case intersections, a sample size of approximately 950 measurements during an eight to twelve day aggregation period is measured.
For the delay a resemblance to the time-dependent stochastic delay model was found for the greater part of the turning movements. The greatest deviation for the delay is measured during congested situations (rush hour periods) and for periods with a very low traffic demand (night time) and can lead up to differences of 20 seconds. The average delay distribution shows more variance than the time-dependent stochastic delay model which appears to be caused by a too small sample size of 7,262 observations at intersection 1 and 12,980 observations at intersection 2 during the three month observation period. Smoothing filters can considerably clarify the trends in the data but lower the level of detail reducing some peaks by 60% and it is advised to use unsmoothed data to avoid the loss of information. The delay of probe vehicles appears to be normally distributed for a single movement indicating that the probes are randomly distributed amongst traffic.

Probe data from consumer GPS navigation devices is beneficial for measurement of delay and turning movements at intersections without fixed sensors or where data from stationary sensors is not easily accessible. A combination with other probe data sources and a growing market share increase the possible applications of the data.

REFERENCES